Machine Learning for Medical Image Analysis & Reconstruction

Stefan Klein
s.klein@erasmusmc.nl
Example: Radiogenomics in glioma

- 1p/19q co-deletion determines survival & treatment response in low-grade glioma
- AIM: predict 1p/19q status from MRI

Image courtesy: Marion Smits
Radiogenomics: predicting genetic mutation status from non-invasive imaging data

Multiparametric Imaging

Image features

- Intensity
- Shape
- Voxel-based

Machine learning

Prediction

Low Grade Gliona

IDH 1/2 mutation

IDH wildtype

1p/19q co-deleted

1p/19q non-co-deleted

Good prognosis

Average prognosis

Bad prognosis

Slide by Sebastian van der Voort
Radiogenomics: predicting genetic mutation status from non-invasive imaging data

- 284 MRI scans (T1w, T2w) from multiple scanners
- Internal cross-validation
- Results:
  - Sensitivity: 0.66 (prediction of co-deleted)
  - Specificity: 0.72 (prediction of non-co-deleted)
  - AUC: 0.76
Radiogenomics: predicting genetic mutation status from non-invasive imaging data

- 129 MRI scans (T1w, T2w) (external public dataset)
- Train on internal data, test on external data
- Results:
  - Sensitivity: .... (prediction of co-deleted)
  - Specificity: .... (prediction of non-co-deleted)
  - AUC: ....
Radiogenomics: predicting genetic mutation status from non-invasive imaging data

- 129 MRI scans (T1w, T2w) (external public dataset)
- Train on internal data, test on external data
- Results:
  - Sensitivity: 0.73 (prediction of co-deleted)
  - Specificity: 0.62 (prediction of non-co-deleted)
  - AUC: 0.74
Ongoing work:

- Collection of 2500 MRI scans
- Both low-grade and high-grade glioma
- Genetic markers: 1p/19q, IDH, MGMT
- T1w pre/post-contrast, T2w, FLAIR, DWI, PWI
- Sources: 4 public and 3 internal datasets
Radiogenomics: predicting genetic mutation status from non-invasive imaging data

• Two best examples as indicated by the algorithm

(\textit{left: mutation; right: no mutation})
Radiogenomics: predicting genetic mutation status from non-invasive imaging data

- Examples from literature (Smits et al., 2016)

  *(left: mutation; right: no mutation)*
AI in medical image analysis: a mechanical engineering perspective
WORC: Workflow for Optimal Radiomics Classification

Online Multiparametric Database
- Lung PET
- Liver CT
- Head Desmoid T1-weighted MR
- Liver T2-weighted MR
- Brain T1-Weighted MR
- Mammogram

Radiomics Platform
- Segmentation
- Classification
- Registration
- Transformation
- Feature Extraction

Predictions
- Genetic Mutations
- Tumor Phenotype
- Therapy Response
- Patient Prognosis
- Dementia Diagnosis
- Quantitative maps

Slide by Martijn Starmans
WORC: Workflow for Optimal Radiomics Classification

Online Multiparametric Database
- Lung PET
- Liver CT
- Head Desmoid T1-weighted MR
- Liver T2-weighted MR
- Brain T1-Weighted MR
- Mammogram

Radiomics Platform

Predictions
- Genetic Mutations
- Tumor Phenotype
- Therapy Response
- Patient Prognosis
- Dementia Diagnosis
- Quantitative maps

Slide by Martijn Starmans
## WORC: Results

<table>
<thead>
<tr>
<th>Label Modality</th>
<th>Liver</th>
<th>Brain</th>
<th>Lipo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malignancy</td>
<td>T2w (FatSat) MR</td>
<td>Mutation T2w + T1w MR</td>
<td>Amplification T1w MR</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Label Modality</th>
<th>GI Tract</th>
<th>Mets</th>
<th>Prostate</th>
<th>Desmoids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgery CT</td>
<td>Growth pattern CT</td>
<td>Risk T2w, ADC, DWI</td>
<td>Phenotype T1w MR</td>
<td></td>
</tr>
</tbody>
</table>
WORC: Results

Reported as **mean** [95% confidence interval]

<table>
<thead>
<tr>
<th></th>
<th>Liver</th>
<th>Brain</th>
<th>Lipo</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td><strong>0.93</strong> [ 0.86, 0.99 ]</td>
<td><strong>0.80</strong> [ 0.74, 0.85 ]</td>
<td><strong>0.84</strong> [ 0.74, 0.93 ]</td>
</tr>
<tr>
<td>F1-score</td>
<td><strong>0.82</strong> [ 0.76; 0.91 ]</td>
<td><strong>0.76</strong> [ 0.71, 0.80 ]</td>
<td><strong>0.76</strong> [ 0.66, 0.85 ]</td>
</tr>
<tr>
<td>Sensitivity</td>
<td><strong>0.74</strong> [ 0.58, 0.89 ]</td>
<td><strong>0.67</strong> [ 0.58, 0.76 ]</td>
<td><strong>0.73</strong> [ 0.59, 0.86 ]</td>
</tr>
<tr>
<td>Specificity</td>
<td><strong>0.92</strong> [ 0.85, 0.98 ]</td>
<td><strong>0.79</strong> [ 0.72, 0.86 ]</td>
<td><strong>0.80</strong> [ 0.67, 0.92 ]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>GI Tract</th>
<th>Mets</th>
<th>Prostate</th>
<th>Desmoids</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td><strong>0.74</strong> [ 0.55, 0.92 ]</td>
<td><strong>0.68</strong> [ 0.57, 0.79 ]</td>
<td><strong>0.74</strong> [ 0.59, 0.91 ]</td>
<td><strong>0.88</strong> [ 0.80, 0.95 ]</td>
</tr>
<tr>
<td>F1-score</td>
<td><strong>0.65</strong> [ 0.48; 0.82 ]</td>
<td><strong>0.63</strong> [ 0.53, 0.73 ]</td>
<td><strong>0.85</strong> [ 0.70, 1.00 ]</td>
<td><strong>0.82</strong> [ 0.74; 0.90 ]</td>
</tr>
<tr>
<td>Sensitivity</td>
<td><strong>0.70</strong> [ 0.43, 0.96 ]</td>
<td><strong>0.65</strong> [ 0.48, 0.81 ]</td>
<td><strong>0.60</strong> [ 0.60, 0.78 ]</td>
<td><strong>0.85</strong> [ 0.73, 0.96 ]</td>
</tr>
<tr>
<td>Specificity</td>
<td><strong>0.64</strong> [ 0.43, 0.85 ]</td>
<td><strong>0.63</strong> [ 0.48, 0.77 ]</td>
<td><strong>0.71</strong> [ 0.42, 0.84 ]</td>
<td><strong>0.80</strong> [ 0.67, 0.92 ]</td>
</tr>
</tbody>
</table>
1. Colorectal Liver Metastases: HGP (Validation)
2. Colorectal Liver Metastases: Survival (Segmentation)
3. Colorectal Liver Metastases: Recurrence (Segmentation)
4. Benignity/Phenotype of Liver Tumors (Validation)
5. Prediction of liver metastasizing (Setup)
6. Prediction of HCC in screening (Segmentation)
7. Prediction of MVI in HCC (Setup)
8. Breast Cancer Radiotherapy Response (Radiomics)
9. Desmoids: Mutations (Radiomics)
10. Desmoids: Differential Diagnosis (Radiomics)
11. Liposomas and Liposarcomas (Radiomics)
12. Mesenterial Fibroses (Radiomics)
13. GIST: Mutations (Radiomics)
14. GIST: Differential Diagnosis (Radiomics)
15. Metastases of Melanoma (Radiomics)
16. Intracranial Calcifications
17. Quantification of birth control effects on HCA (Segmentation)
18. Classification of COPD patients on Video Thermography (Pilot)
19. Low Grade Glioma 1p/19q mutation (Sebastian)
20. Prostate (Jose)
The (often not shown) bad results

WORC does not always perform well immediately. Predicting:
- Therapy response on mammograms of 38 breast cancer patients.
- BRAF mutation on lung CT scans of 50 melanoma patients.
- Three mutations on T2 MR scans of 60 desmoid patients.
- Four mutations on CT scans of 120 GIST patients.

<table>
<thead>
<tr>
<th></th>
<th>Breast</th>
<th>Mela</th>
<th>DM ($\mu$)</th>
<th>GIST ($\mu$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>[0.43, 0.88]</td>
<td>[0.26, 0.59]</td>
<td>[0.44, 0.74]</td>
<td>[0.38, 0.66]</td>
</tr>
<tr>
<td>F1-score</td>
<td>[0.44; 0.75]</td>
<td>[0.32, 0.58]</td>
<td>[0.50, 0.71]</td>
<td>[0.60, 0.73]</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>[0.05, 0.55]</td>
<td>[0.16, 0.89]</td>
<td>[0.07, 0.39]</td>
<td>[0.28, 0.62]</td>
</tr>
<tr>
<td>Specificity</td>
<td>[0.67, 0.99]</td>
<td>[0.34, 0.73]</td>
<td>[0.73, 0.98]</td>
<td>[0.45, 0.74]</td>
</tr>
</tbody>
</table>

Common issues: small datasets (< 50 patients), low prevalence (10-30 example per class), high imbalance and no previous relation between imaging and labels.
Machine learning needs data!
Infrastructural challenges

- Data collection
- Data anonymisation
- **Data clean-up & structuring**
- Data storage
- Data sharing
- Data inspection & annotation
- Data processing & analysis
- Data integration

*traceable & reproducible*
Radiogenomics: predicting genetic mutation status from non-invasive imaging data

Ongoing work:
- Collection of 2500 MRI scans
- Both low-grade and high-grade glioma
- Genetic markers: 1p/19q, IDH, MGMT
- T1w pre/post-contrast, T2w, FLAIR, DWI, PWI
- Sources: 4 public and 3 internal datasets
Problem: sorting MRI scan types!

(a) T1w  (b) T1wC  (c) T2w  (d) PDw

(e) T2w-FLAIR  (f) DWI  (g) DWI-D (ADC)
Idea: use machine learning for MRI scan type recognition!

- Train a convolutional neural network (CNN) on 7153 scans from 665 patients
- Test on 1700 scans from 207 patients

![Fig. 2: Architecture of the CNN](image-url)
**DeepDicomSort**

**Evaluation results on test set:**

<table>
<thead>
<tr>
<th>Scan type</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>98.3%</td>
</tr>
<tr>
<td>T1w</td>
<td>99.1%</td>
</tr>
<tr>
<td>T1w post-contrast</td>
<td>98.5%</td>
</tr>
<tr>
<td>T2w</td>
<td>99.7%</td>
</tr>
<tr>
<td>PDw</td>
<td>99.9%</td>
</tr>
<tr>
<td>T2w-FLAIR</td>
<td>95.9%</td>
</tr>
<tr>
<td>DWI</td>
<td>98.0%</td>
</tr>
<tr>
<td>DWI-derived</td>
<td>96.4%</td>
</tr>
</tbody>
</table>
DeepDicomSort

(a) T1w
(b) T1wC
(c) T2w
(d) PDw

(e) T2w-FLAIR
(f) DWI
(g) DWI-D (ADC)
Machine learning for image reconstruction
Many existing advanced reconstruction algorithms

- Undersampled k-space
- Parallel imaging
- Motion compensation

http://mriquestions.com/what-is-k-space.html
Many existing advanced reconstruction algorithms:

- Undersampled k-space
- Parallel imaging
- Motion compensation
- …
Machine learning for image reconstruction

Images by Zhu et al.
Machine learning for image reconstruction

Images by Zhu et al.
Machine learning for image reconstruction

Images by Zhu et al.
Machine learning for image reconstruction

Images by Zhu et al.
APIR-Net: Autocalibrated Parallel Imaging Reconstruction using a Neural Network

Chaoping Zhang, Florian Dubost, Marleen de Bruijne, Stefan Klein, and Dirk H. J. Poot
c.zhang@erasmusmc.nl
APIR-Net
Autocalibrated Parallel Imaging Reconstruction using a Neural Network

MRI scanner
Multi-channel coil
Image

Acquisition

Subsampled k-space

Interpolated k-space

Improved noise resilience

Linear fitting (least squares)

Nonlinear fitting

iFFT

APIR-Net

N: Sampled  O: Unsampling  #: Computed
To exploit the redundancy among the multiple highly correlated channels in the receive coil, in APIR-Net
1. the number of feature maps decreases as the depth of the encoder increases;
2. the size of each feature map remains unchanged.
APIR-Net
Autocalibrated Parallel Imaging Reconstruction using a Neural Network

APIR-Net achieves
1. reduced noise amplification
2. reduced artifacts
Current work: Machine learning for quantitative MRI image reconstruction

B-Q MINDED: a new H2020 ITN project on q-MRI

B-Q MINDED is an international research project on quantitative magnetic resonance imaging, coordinated by Prof. Dr. J. Sijbers.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 764513.
Current work:
Machine learning for quantitative MRI image reconstruction
Conclusion

Numerous opportunities for use of machine learning in radiology:

- Diagnosis / prognosis
- Disease phenotyping
- Molecular subtyping
- Scan type recognition
- Image reconstruction
- Image quantification
- …
Machine Learning for Medical Image Analysis & Reconstruction

Stefan Klein
s.klein@erasmusmc.nl
Prediction of dementia

Alzheimer’s Disease
<or>
Mild Cognitive Impairment (stable/progressive)
<or>
Normal

Slide by Esther Bron
Segmentation with deep learning

White matter tracts segmentation with V-NET trained on $N=7000$ images, tested on $N=1000$ images.